(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平8－335144

(43)公開日　平成８年(1996)12月17日

(51)Int.Cl.[6]		識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ０６Ｆ	3/06	３０４		Ｇ０６Ｆ	3/06	３０４Ｂ
	11/20	３１０			11/20	３１０Ｂ
	12/16	３１０	7623－5B		12/16	３１０Ｑ
	13/14	３１０	7368－5E		13/14	３１０Ｆ

審査請求　未請求　請求項の数５　ＯＬ　（全 17 頁）

(21)出願番号　特願平7－139781

(22)出願日　平成７年(1995)６月７日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(72)発明者　松本　佳子
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(72)発明者　村岡　健司
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内

(74)代理人　弁理士　筒井　大和

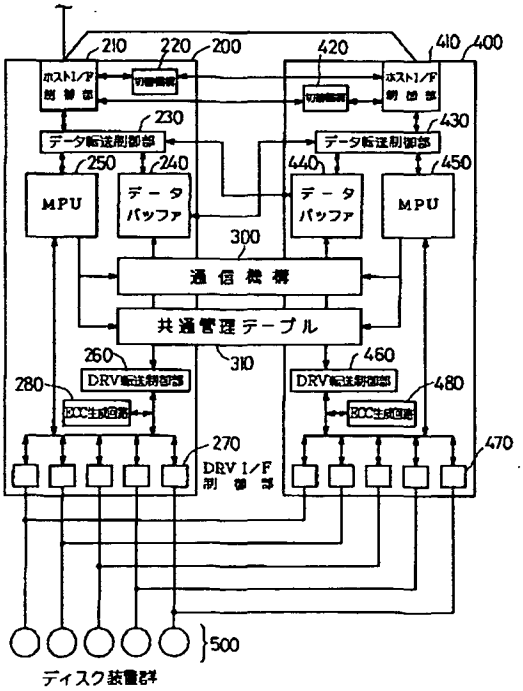
(54)【発明の名称】　外部記憶装置

(57)【要約】

【目的】　冗長構成の複数の記憶制御装置に負荷を分散させて、信頼性および性能を向上させるとともに、無停止保守を実現する。

【構成】　ディスク装置を制御する冗長構成の複数のディスクドライブ制御装置２００および４００を上位装置に対して同一ＳＣＳＩＩＤで接続するとともに、通信機構３００および共通管理テーブル３１０を介在させて相互の稼働状態の監視および負荷分散情報の設定を行い、正常時には、複数のディスクドライブ制御装置２００および４００の同時稼働による負荷分散によって高性能を実現し、障害や保守作業時には、切替機構２２０および４２０によって障害側の切り離しによる縮退および復旧時の切り替え操作を実行して、無停止稼働および無停止保守を実現する。

図２



【特許請求の範囲】

【請求項１】　上位装置との間で授受されるデータが格納される記憶装置と、前記記憶装置と前記上位装置との間に介在し、前記上位装置と前記記憶装置との間における前記データの授受を制御する複数の記憶制御装置とを含む外部記憶装置であって、

複数の前記記憶制御装置が前記上位装置からみて等価に見えるように当該記憶制御装置を前記上位装置に接続するインターフェイス手段と、

個々の前記記憶制御装置に設けられ、他の前記記憶制御装置における障害または切替指令の有無を監視する監視手段と、

個々の前記記憶制御装置に設けられ、いずれの前記記憶制御装置が前記上位装置との間における前記データの授受の制御を行うかを切り替える切替手段と、

前記記憶制御装置の相互間における情報の伝達を行う情報伝達手段と、

前記上位装置からの入出力要求に起因する負荷を複数の前記記憶制御装置間にて分担させる負荷分散手段と、を備えたことを特徴とする外部記憶装置。

【請求項２】　請求項１記載の外部記憶装置において、

複数の前記記憶制御装置の各々に設けられ、前記上位装置との間で授受される前記データを一時的に格納するデータバッファと、

前記上位装置からの書き込み要求時、複数の前記データバッファの各々に対して書き込み要求データを選択的または多重に書き込むとともに、前記書き込み要求データの前記データバッファに対する書き込み完了時点で前記上位装置に対して書き込み完了を報告し、前記上位装置からの入出力要求とは非同期に前記データバッファから前記記憶装置へ前記書き込み要求データを反映させるライトアフタ処理、および前記書き込み要求データの前記記憶装置に対する書き込み完了時点で前記上位装置に対して書き込み完了を報告するライトスルー処理を選択的に実行可能なデータ転送制御手段と、を備えたことを特徴とする外部記憶装置。

【請求項３】　請求項１または２記載の外部記憶装置において、

複数の前記記憶制御装置から共通にアクセス可能にされ、個々の前記記憶制御装置が健全か否かを識別するための第１の管理情報、前記ライトアフタ処理および前記ライトスルー処理の何れを実行するかを指定する第２の管理情報、複数の前記記憶制御装置の何れが前記上位装置からの入出力要求を受け付けるかを指定する第３の管理情報、複数の前記記憶制御装置の各々における前記負荷の分担を指定する第４の管理情報の少なくとも一つが格納される管理情報記憶手段と、

障害の発生または外部からの切替指令を契機として、前記障害が発生したか、または外部から指令された前記記憶制御装置を切り離すとともに、残りの前記記憶制御装置によって前記上位装置との間における前記データの授受を継続する縮退運転を行う操作、および切り離されていた前記記憶制御装置を冗長構成に復帰させる操作を行う制御論理と、を備えたことを特徴とする外部記憶装置。

【請求項４】　請求項２記載の外部記憶装置において、

前記データ転送制御手段は、個々の前記記憶制御装置の各々に設けられた前記データバッファの各々に対する前記書き込み要求データの選択的な書き込み操作の停止および再開を行う制御論理を備えたことを特徴とする外部記憶装置。

【請求項５】　請求項４記載の外部記憶装置において、

複数の前記記憶制御装置の中の少なくとも一つを選択的に停止させて縮退運転を行うとともに、停止された前記記憶制御装置に対応するデータバッファの保守または前記記憶制御装置を制御するマイクロプログラムの保守を実行することを特徴とする外部記憶装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、外部記憶装置に関し、特に、上位装置からの情報の入出力要求を制御する入出力制御装置を多重に備えた冗長構成の外部記憶サブシステム等に適用して有効な技術に関する。

【０００２】

【従来の技術】コンピュータシステムを構成する外部記憶装置においては、記憶媒体を備えた記憶装置と上位装置との間に介在して両者間の情報の授受を制御する記憶制御装置が冗長構成でない場合、記憶制御装置に障害が発生するとサブシステムは停止を余儀なくされ、この間に復旧作業がおこなわれる。そして、この復旧作業が終了すると、記憶制御装置が再起動され、あるいは、サブシステムが再起動されそれまで中断していた業務が再開される。

【０００３】また最近では、コンピュータシステムを用いる様々な情報処理業務において２４時間稼働の運用形態が増加しており、外部記憶サブシステムにも連続運転が要求されている。このため、たとえば、特開平３－２０６５２９号公報に記載されているように、一方の記憶制御装置が運転中、他系の記憶制御装置は停止してスタンバイ状態をとるという、記憶制御装置に関して冗長な構成をとり、記憶制御装置の障害時、他系のスタンバイ状態の記憶制御装置に切り替わることにより、システムを継続運転可能にしようとする技術が知られている。

【０００４】

【発明が解決しようとする課題】しかし、上述の従来技術においては、障害時の連続運転は可能であるが、２台の記憶制御装置を備えているにもかかわらず、実際に稼働するのはいずれか１台のみであり性能的には１台の時となんら変わらなかった。すなわち冗長な他系の記憶制御装置はあくまでホットスタンバイ用であり、障害の記

憶制御装置の単なる代替でしかなかった。

【０００５】また、近年では、システムへの要求も様々であり、上位装置からも複数の経路より、同一の記憶装置へ、または別々の記憶装置へとアクセス要求が発行されるような様々な接続形態があり、従来のような記憶制御装置の単なる冗長構成では、多様なユーザの要請に合わせてシステムを構築することが困難であった。

【０００６】また、従来において、安価なシステムでは１つのボード内に記憶制御装置とデータバッファが搭載された構成となっており、記憶制御装置内のデータバッファの増設等の保守管理を行う場合、データバッファのみの切り放しが不可能なため、システムを一旦停止させた状態でデータバッファを増設し、増設作業の終了後、記憶制御装置やシステムを再起動させ、それまで中断していた業務を再開する、という手順を踏む必要があり、上位装置からの入出力要求（Ｉ／Ｏ）を処理しながら増設等の保守管理作業を遂行することは不可能であった。

【０００７】本発明の目的は、冗長構成の複数の記憶制御装置に負荷を分散させることにより、信頼性および性能を向上させることが可能な外部記憶装置を提供することにある。

【０００８】本発明の他の目的は、上位装置の側に記憶制御装置の冗長構成を意識させることなく、記憶制御装置の多重化による信頼性の向上、さらには記憶制御装置の多様な制御動作を実現することが可能な外部記憶装置を提供することにある。

【０００９】本発明のさらに他の目的は、稼働を停止させることなく、冗長構成の複数の記憶制御装置におけるハードウェアおよびソフトウェア等の保守管理作業を簡便に遂行することが可能な外部記憶装置を提供することにある。

【００１０】本発明のさらに他の目的は、単一のボード上に記憶制御装置およびデータバッファを搭載した構成の保守管理作業を、稼働中に実行することが可能な外部記憶装置を提供することにある。

【００１１】

【課題を解決するための手段】本発明の外部記憶装置は、複数の記憶制御装置が上位装置からみて同一に見えるように記憶制御装置を上位装置に接続するインターフェイス手段と、複数の記憶制御装置間で他の記憶制御装置を監視する監視手段と、記憶制御装置間での情報の伝達が可能な通信手段と、上位装置からの要求を受領している記憶制御装置を切り替える切替手段と、１つの記憶制御装置が受領した上位装置からの入出力要求と、それに付随する処理を複数の記憶制御装置にて負荷分散する負荷分散手段とを含む構成としたものである。

【００１２】また、個々の記憶制御装置に備えられ、記憶制御装置と同様の冗長構成をとることによって上位装置からの書き込みデータを一旦格納するデータバッファと、上位装置からの書き込みデータをデータバッファに格納した時点で上位装置へ終了を報告し、上位装置との要求とは非同期にデータバッファから記憶装置に書き込むとともに、冗長構成の複数のデータバッファのすべてに書き込むか選択的に書き込むかを制御するデータ転送手段とを含む構成としたものである。

【００１３】また、複数の記憶制御装置から共通にアクセス可能にされ、個々の記憶制御装置が健全か否かを識別するための第１の管理情報、ライトアフタ処理およびライトスルー処理の何れを実行するかを指定する第２の管理情報、複数の記憶制御装置の何れが上位装置からの入出力要求を受け付けるかを指定する第３の管理情報、複数の記憶制御装置の各々における負荷の分担を指定する第４の管理情報、の少なくとも一つが格納される管理情報記憶手段を含む構成としたものである。

【００１４】

【作用】本発明の外部記憶装置では、たとえば、一対の第１および第２の記憶制御装置を含む冗長構成であるとき、この第１および第２の記憶制御装置は上位装置と、たとえばＳＣＳＩインターフェースによりデージーチェーン接続され、同一ＳＣＳＩＩＤでアクセスされる。たとえば上位装置からの入出力要求を固定的に第１の記憶制御装置が受領している場合、当該入出力要求の処理に伴う負荷は、複数の記憶制御装置の各々における負荷の分担を指定する第４の管理情報に基づいて他の第２の記憶制御装置に分散され、冗長構成による信頼性の向上と第１および第２の記憶制御装置の並行稼働による入出力処理の処理能力の向上が図れる。

【００１５】また、たとえば上位装置からの入出力要求を固定的に第１の記憶制御装置が受領している場合、第１の記憶制御装置に障害が発生した時、第１の管理情報および監視手段により障害を検出し、切替手段によって要求を受領する記憶制御装置を第２の記憶制御装置に切り替えることにより、上位装置は障害後も同一のＳＣＳＩＩＤに対してＩ／Ｏ要求を発行すればよく、切り替えに対してなんら意識する必要はない。その後、障害となった記憶制御装置を切り放し、縮退運転に入る。部品やマイクロプログラムの交換等の保守作業終了後に第１の記憶制御装置を復旧し、元の冗長構成に復元される。

【００１６】また、各記憶制御装置内にデータバッファを持ち、上位装置からの書き込みデータを第１の記憶制御装置が受領し、ライトアフタ処理を実行している場合、第１の記憶制御装置に障害が発生したことを監視手段により検出し、切替手段により、第１の記憶制御装置から第２の記憶制御装置に入出力要求を受領する記憶制御装置を切り替えた時、同時に、複数のデータバッファに対する多重なデータ書き込み処理から、稼働中の記憶制御装置に備えられたデータバッファに選択的にデータを書き込む処理に切り替える。

【００１７】この時、ライトアフタ処理を実行するかライトスルー処理を実行するかを選択する。この選択は、

管理情報記憶手段の第２の管理情報をユーザが設定することにより可能である。すなわち、ユーザのデータ信頼性に対する要求が高い時は、ライトスルーモードに設定し、信頼性よりは性能を要求する場合は、ライトアフタモードに設定する。

【００１８】第２の記憶制御装置の復旧後、選択的な書き込みか多重書き込みに切り替えることによりデータバッファに多重に書き込む操作に切り替え、冗長構成に復元できる。

【００１９】上位装置からの入出力要求を第１の記憶制御装置が受領し、上位装置からの書き込み要求に関しては、第１の記憶制御装置のデータバッファと第２の記憶制御装置のデータバッファに２重書きを行い、ライトアフタ処理を行なっている場合、多重に書き込む処理から選択的に書き込む処理に切り替えて第２の記憶制御装置を切り放して縮退させ、データバッファの増設やマイクロプログラムの交換等の保守を行なった後、元の冗長構成に復旧させる。その後、記憶制御装置間での情報の伝達が可能な通信手段用いて第２の記憶制御装置は第１の記憶制御装置にデータバッファの増設等の保守作業の完了を通知し、通知後、切替手段を用い、上位装置からの要求の受領を自装置に切り替える。

【００２０】一方、通知を受けた第１の記憶制御装置は自装置を縮退させ、データバッファの増設等の保守を行ない復旧させる。復旧後、情報の伝達が可能な通信手段を用いて第１の記憶制御装置は第２の記憶制御装置にデータバッファの増設等の保守完了を通知する。これを契機に、単一のデータバッファに対する選択的なデータ書き込みから、複数のデータバッファに対する多重書き込み処理に切り替える。これにより、一対の第１の記憶制御装置および第２の記憶制御装置のデータバッファの増設やマイクロプログラムの交換等の保守管理業務を、上位装置との間における入出力処理を継続しながら可能となる。

【００２１】また、本発明によれば、管理情報記憶手段に設定されている複数の前記記憶制御装置の何れが前記上位装置からの入出力要求を受け付けるかを指定する第３の管理情報を参照することにより、第１の記憶制御装置および第２の記憶制御装置は、上位装置からの要求をいずれが受領するかを判断することが可能である。これにより、たとえば、第１の記憶制御装置および第２の記憶制御装置のどちらか一方のみが上位装置からの要求を受け付けることに限らず、両方の記憶制御装置にて入出力要求を受け付けて処理することも可能である。また、第３の管理情報をユーザにて随意に設定することにより、上位装置からの要求を受ける記憶制御装置をユーザから任意に指定することが可能となる。

【００２２】

【実施例】以下、本発明の実施例を図面を参照しながら詳細に説明する。

【００２３】図１は本発明の一実施例である外部記憶装置を含む計算機システムの一例を示す概念図である。本実施例の計算機システムは、中央処理装置である上位装置１００と、ディスクドライブ制御装置２００、ディスクドライブ制御装置４００と、ディスク装置５００とを含んでいる。ディスクドライブ制御装置２００とディスクドライブ制御装置４００は上位装置１００とＳＣＳＩインターフェースのデージーチェーンで接続され、ディスクドライブ制御装置２００、４００は同一のＳＣＳＩＩＤが設定され、冗長構成をとっている。そして本実施例の場合、ディスクドライブ制御装置２００は、上位装置１００からの要求を受領し、要求に付随する処理をディスクドライブ制御装置２００と冗長なディスクドライブ制御装置４００とで実行し、ディスク装置５００を制御する。

【００２４】図２は、ディスクドライブ制御装置２００および４００の内部構成の一例を示すブロック図である。なお、ディスクドライブ制御装置２００とディスクドライブ制御装置４００の内部構成は同一であるため、ディスクドライブ制御装置２００を例に説明し、ディスクドライブ制御装置４００の側については対応する部位の符号の下２桁を同一にして説明は割愛する。

【００２５】マイクロプロセッサユニット２５０（以下ＭＰＵと称す）は、ランダムアクセスメモリ（ＲＡＭ、図示せず）を逐次デコードしながら実行し、ディスクドライブ制御装置２００の全体を制御している。

【００２６】ホストＩ／Ｆ制御部２１０は、上位装置１００とのプロトコル制御を行なっている。ＤＲＶＩ／Ｆ制御部２７０は各ドライブとのプロトコル制御を行なっている。データバッファ２４０は、ホストＩ／Ｆ制御部２１０とＤＲＶＩ／Ｆ制御部２７０のデータ転送時に用いられるものである。このメモリは揮発メモリでもよいし不揮発メモリでもよい。本実施例は揮発メモリでデータバッファ２４０を構築した場合を例に記述する。

【００２７】切替機構２２０は、各ディスクドライブ制御装置２００およびディスクドライブ制御装置４００のホストＩ／Ｆ制御部２１０およびホストＩ／Ｆ制御部４１０に対し、上位装置１００からのＩ／Ｏを受領するホストＩ／Ｆ制御部を切り替えるためのものである。この実施例では、ホストＩ／Ｆ制御部２１０が受領しているものとする。データ転送制御部２３０は上位装置１００とデータバッファ２４０とのデータ転送を制御している。このデータ転送制御部２３０は上位装置１００からのライトデータをデータバッファ２４０とデータバッファ４４０の２面に２重書きするか、データバッファ２４０のみの１重書きするかの両方の機能を備えている。また、ＭＰＵ２５０からの指示により１重書きか２重書きかを切り替えることが可能である。

【００２８】ＤＲＶ転送制御部２６０はデータバッファ２４０とディスク装置５００との間のデータ転送を制御

する。

【0029】通信機構300は、MPU250と、MPU450間での情報の伝達をするための機構である。この通信機構300はMPU250とMPU450間における双方向の伝達を可能としている。

【0030】共通管理テーブル310は、MPU250とMPU450の双方から参照／更新が可能な管理テーブルである。

【0031】本実施例では、上位装置100からの論理データを複数のディスク装置500へ分散させて格納する、アレイ構成によるドライブ格納方式を例に採って説明する。

【0032】ECC生成回路280は、上位装置100より送られてきたデータに対して冗長データを生成する機能を有し、この機能はデータの復元にも用いることができる。冗長データを付加する単位は、上位から送られてきた1論理データ単位でもよいし複数の論理データ単位に対してでもよい。本実施例は、4つの論理データに対し冗長データを付加し、この冗長データを格納するドライブを固定しないRAID5方式において記述する。

【0033】次に、図3を参照して共通管理テーブル310の構成の一例について説明する。監視情報320は各ディスクドライブ制御装置200／400が正常に動作しているかどうかをチェックするのに用いられる。監視情報A321はディスクドライブ制御装置200のMPU250が正常に動作可能と判断された時、一定間隔にて情報を設定する。また、MPU250が正常に動作不能と判断した時、異常を示す情報を設定する。なお、ディスクドライブ制御装置400のMPU450も、MPU250と同様に情報を監視情報B322に設定する。

【0034】データ転送モード情報330は、システムの縮退状態時に上位装置100からのライトデータ書き込み要求に対する終了報告契機を指示する。すなわち、この情報がデータバッファ240、またはデータバッファ440に対する書きこみ完了時点で上位装置100に終了を報告するか（以下ライトアフタモードと称す）、データバッファ240からディスク装置500にまで書き込んだ時点で終了報告するか（以下ライトスルーモードと称す）を判断するための情報である。

【0035】ホストI／O受信情報340は2つのディスクドライブ制御装置200／400の内、I／Oを受信するディスクドライブ制御装置の指示情報が示されている。本実施例では、ディスクドライブ制御装置200がホストI／O受信側に設定されているものとして説明する。

【0036】負荷分散情報350は、上位装置からのI／Oに伴う処理を、2つのディスクドライブ制御装置200／400間で負荷分散するための情報である。負荷分散の方法は各ディスクドライブ制御装置にアクセス対象とするディスク装置を分割してもよいし、上位装置100からのI／O要求の処理と、上位装置100からのI／O要求とは非同期のデータバッファからディスク装置500へライトデータを格納する処理とに分担してもよい。または処理しなければならない事柄を全て負荷分散情報の中に書き込み、2つのMPU間で競争論理とし、MPUとして空きがあるほうが処理を実行するという方法でもかまわない。

【0037】本実施例では上位装置100からのI／O要求の処理と、上位装置100からのI／O要求とは非同期にデータバッファ240／440からディスク装置500へライトデータを格納する処理とに分担する方式について説明する。よって、本実施例では、負荷分散情報350にはデータバッファ240／440に格納されたライトデータの情報が入っているものとする。

【0038】次に本実施例における計算機システムでの、上位装置100からディスク装置500に対するデータの書き込み処理および読み込み処理について説明する。

【0039】ディスクドライブ制御装置200は、通常、上位装置100からの書き込み要求時、ホストI／F制御部210により、書き込み論理データを受領し、データ転送制御部230にてデータバッファ240とデータバッファ440に2重に格納し、共通管理テーブル310の負荷分散情報350に格納情報を設定し、この時点で上位装置100に終了を報告する。MPU450は、逐次、負荷分散情報350を参照し、格納情報があれば、当該ライトデータと同一アドレスの、既にドライブに格納されているデータ（以下旧データと称す）と、当該ライトデータに対応するパリティデータをDRVI／F制御部470とDRV転送制御部460によりディスク装置500から読み出し、ECC生成回路480にてライトデータと旧データとパリティデータにて、ライトデータに対応したパリティデータ（以下新パリティデータと称す）を生成する。生成された新パリティデータとライトデータをDRVI／F制御部470とDRV転送制御部460によりディスク装置500に書き込むことにより、ライトデータをディスク装置500に格納する。この処理は上位装置100からのI／O要求とは非同期に行なわれる。また、ライトデータを格納するために行なわれる、旧データ／旧パリティデータの読み出し処理及び新パリティ生成処理、新パリティデータ格納処理はRAID5におけるライトペナルティと呼ばれている。

【0040】このように、上位装置100からのライトデータの格納要求は、複数のディスク装置500をディスクアレイ装置として機能させる場合において、非常に負荷が高い処理である。この処理を2つのディスクドライブ制御装置200、400にて役割分担し実行することは、1台のディスクドライブ制御装置だけで実行する

より、効率がよくシステムとしての性能向上につながる。特に最近の市場動向としては、安価なプロセッサを搭載し、システム全体のコストを低減させることが、高性能、高信頼性とともに、非常に大切な要素となっている。よって、ライトペナルティ処理においては、ドライブへのアクセスが多数発生することも性能劣化につながるが、それ以前にそれを制御するプロセッサのマイクロプログラムの走行時間が長いために、システムとしてプロセッサネックになることも多い。この時、本実施例のように、２台のディスクドライブ制御装置２００および４００にて処理を行なうことで２倍近くの性能を出すことができる。

【００４１】次に上位装置１００からの読み込み要求時、ＭＰＵ２５０は、ＤＲＶＩ／Ｆ制御部２７０とＤＲＶ転送制御部２６０により物理ドライブ（ディスク装置５００）よりデータの読み込みを開始し、上位装置１００に転送する。また、この時、上位装置１００からのリード要求アドレスが連続していた時、ディスクドライブ制御装置４００がシーケンシャルリード処理だと判断し、上位装置１００からのリード要求アドレスに続くあるデータ量を上位装置１００からのＩ／Ｏとは非同期にデータバッファ２４０および４４０に読み出す処理を行っても良い。こうすることにより、次に上位装置からＩ／Ｏ要求があった時、対象となるデータがすでにデータバッファ２４０／４４０に格納されており、時間の掛かるディスク装置５００へのアクセスを生じることなくデータを転送することができ、全体としての性能向上につながる。

【００４２】以上のように、冗長構成でありながら、冗長な部分（本実施例では、ディスクドライブ制御装置４００）を障害発生時の切り替え用として単にスタンバイさせておくのではなく、処理の一部を実行させることにより、信頼性だけでなく性能の向上にもつながる。

【００４３】次に本実施例において、２台のディスクドライブ制御装置２００／４００が処理を実行しながら、障害時の自動切り替えおよび復旧を実行する動作について説明する。まず、自動的に障害を検出する監視手続きについて説明する。

【００４４】ＭＰＵ２５０、４５０はディスクドライブ制御装置２００、４００を制御しながら、一定時間が経過する度に、ＭＰＵ２５０は監視情報３２１に、ＭＰＵ４５０は監視情報３２２に正常であることを示す情報（以下、正常情報と称す）を設定する。但し、一定時間毎に設定していることを示すために、この情報には逐次変化する情報を設定する。たとえば、１つずつ加算されるような情報である。また、各ＭＰＵ２５０，４５０が、当該ディスクドライブ制御装置２００、４００にて正常に動作が不可能と判断したと、たとえば、ＭＰＵからデータバッファがアクセス不可能となった時、監視情報に障害であることを示す情報（以下これを障害情報と称す）を設定する。以下、図４のフローチャートにより前述の監視手続きの一例を説明する。

【００４５】ここでは、ディスクドライブ制御装置４００のＭＰＵ４５０が他系のディスクドライブ制御装置２００の監視を行う動作を例に採り説明する。

【００４６】まずＭＰＵ２５０はステップ６００にて一定時間が経過したかを判断する。一定時間が経過していなければ、ステップ６０８へ進み、ディスクドライブ制御装置２００が正常と判断する。

【００４７】一定時間が経過していれば、ステップ６０１へ進み、ＭＰＵ４５０が正常であることを示す正常情報を設定する。そしてステップ６０２へ進み、ディスクドライブ制御装置２００の監視情報３２２を参照する。この情報が正常か否かを判断し、ステップ６０３にて正常だと判断したら、ステップ６０４に進む。障害であると判断したら、ステップ６０７に進み、ディスクドライブ制御装置２００は障害であると判断する。

【００４８】正常情報の時、ステップ６０５に進み、この正常情報に以前から変更があったかどうかをステップ６０５にて判断する。すなわち、ＭＰＵ２５０がマイクロプログラムの障害等により、監視情報を設定不可能に陥っている可能性がある。このような障害を、このステップ６０５のチェックにて判断する。変更があれば、ステップ６０８へ進み、正常と判断する。変更が無かったとき、ステップ６０６に進み、一定時間よりも長いマージンの時間が経過しているか否かを判断する。その結果、経過していれば、ステップ６０７へ進み、障害と判断し、経過していなければ、ステップ６０８へ進み、正常と判断する。以上のような監視手続きによれば、ハードウェアの障害も、マイクロプログラムの障害も両方同時に検出が可能である。

【００４９】次に、図５のフローチャートを参照してディスクドライブ制御装置４００が他系のディスクドライブ制御装置２００の障害を認識して切り替わる処理の一例を説明する。

【００５０】まずＭＰＵ４５０はステップ７００にて逐次、負荷分散情報３５０を参照している。その結果、ステップ７０１にてデータバッファ２４０、４４０内に上位装置１００からのライトデータが存在しなければ、ステップ７０４に進む。存在すれば、ステップ７０２に進み、データバッファ４４０のライトデータに対応するパリティを生成するため、当該ライトデータに対応する旧データと旧パリティデータをディスク装置５００から読み出し、ＥＣＣ生成回路４８０にて新パリティデータを生成する。その後、ステップ７０３に進み、ライトデータと新パリティデータをＤＲＶ転送制御部４６０および、ＤＲＶＩ／Ｆ制御部４７０によりディスク装置５００に格納する。次にステップ７０４にて、図４のステップ６００以降の監視手続きによりディスクドライブ制御装置２００の障害をチェックする。その結果、正常なら

ば、ステップ７００に進み、処理を続ける。切り替えが必要と判断したら、ステップ７１０に進み、切り替え手続きを用いて、上位装置１００からのＩ／Ｏの受信をディスクドライブ制御装置２００からディスクドライブ制御装置４００に切り替える。そして、ディスクドライブ制御装置２００が行なっていた上位装置１００からのＩ／Ｏ処理をステップ７２０にてディスクドライブ制御装置４００が代替して行なう。

【００５１】次に、図６のフローチャートにて切り替え手続きの一例を説明する。

【００５２】まずステップ７１１にて、データ転送制御部４３０に対し、上位装置１００からのライトデータ受領時、当該データをデータバッファ４４０へ１重に書き込むことを指示する。すなわち、ディスクドライブ制御装置２００に障害が発生したため、ディスクドライブ制御装置２００を縮退させて切り放し、障害発生した部位を交換し、復旧するまでの間、データバッファはディスクドライブ制御装置４００にしか存在しないため、正常な冗長構成時のような２重書きはできない。

【００５３】そして、ステップ７１２にて、切替機構４２０にて上位装置１００からのＩ／Ｏ要求をホストＩ／Ｆ制御部２１０から、ホストＩ／Ｆ制御部４１０に切り替えるよう指示をする。これにより、ホストＩ／Ｆ制御部２１０は上位装置１００からの要求を受け付けなくなり、また、ホストＩ／Ｆ制御部４１０は上位装置１００からの要求を受け付けるようになり、実質的にはディスクドライブ制御装置が切り替わることになるが、本実施例ではＳＣＳＩＩＤが同一なため、上位装置１００はＩ／Ｏを切り替える以前と同様のＳＣＳＩＩＤに発行すればよく、受領側のディスクドライブ制御装置が切り替わったことを知る必要が全くない。

【００５４】次に図７のフローチャートを用いて、切り替わった後、ディスクドライブ制御装置４００にてＩ／Ｏを実行する手順の一例を説明する。

【００５５】ステップ７２１にて上位装置１００よりＩ／Ｏ処理を受信すると、ステップ７２２に進み、リード要求かライト要求かを判断する。リード要求の時、ステップ７２９に進み、当該リード要求に対応するディスク装置５００よりデータバッファ４４０に対象データを読み込む。ステップ７３０に進み、データバッファ４４０から上位装置１００へデータを転送し、ステップ７２８にて上位装置１００に対し、終了を報告する。

【００５６】ライト要求の時、ステップ７２３に進み、データバッファ４４０にライトデータを格納する。さらに、ステップ７２４に進み、データ転送モード情報３３０を参照し、ステップ７２５でライトスルーモードか否かを判定する。その結果、ライトアフタモードの時、すなわちデータバッファ４４０に格納した時点で上位装置１００に対して終了を報告するモードの時は、ステップ７２８に進み、終了を報告し、その後、非同期にデータバッファ４４０からディスク装置５００に格納する。ライトスルーモードの時は、ステップ７２６に進み、ライトデータに対するパリティデータを作成し、ステップ７２７にてライトデータと新パリティデータをディスク装置５００に格納し、ステップ７２８にて終了を報告する。さらに、この後、図５のフローチャートにおけるステップ７００からステップ７０３を実行し、切り替え以前の処理も実行する。

【００５７】このように、本実施例によれば、上位装置１００からの指示なしに、上位装置１００になんの意識もさせずに、ディスクドライブ制御装置２００／４００にて、自動的に相互間の切り替え動作および処理の続行が可能である。

【００５８】次に、ディスクドライブ制御装置２００が復旧し、元の冗長構成に戻る時の方法の一例を説明する。

【００５９】まず、図８に示されるフローチャートにて、ディスクドライブ制御装置２００の側の復旧動作の一例について説明する。ステップ８１０で、通信機構３００にてディスクドライブ制御装置４００に対して復旧が完了したことを通知する。その後、ディスクドライブ制御装置２００が冗長なディスクドライブ制御装置となり、以前と立場が入れ替わる。ステップ８１１にて、以前、ディスクドライブ制御装置４００が行なっていた非同期のデステージ処理（図５のステップ７００～７０５）を行なう。

【００６０】さらに、図９に示されるフローチャートにて、通知を受けた側のディスクドライブ制御装置４００の動作の一例を説明する。

【００６１】ステップ８２０にて通信機構３００にてディスクドライブ制御装置２００の復旧完了を認識すると、ステップ８２１でデータ転送制御部４３０にデータバッファ２４０と４４０への２重書きを指示し、ステップ８２１にて上位装置１００からのＩ／Ｏ処理のみを実行する。このように、上位装置１００からのＩ／Ｏ要求を受けながら、元の冗長な構成に復旧することが可能であり、さらに２つのディスクドライブ制御装置２００／４００で処理を負荷分散することにより、性能の向上も図れる。

【００６２】次に、ディスクドライブ制御装置２００／４００の稼働中におけるデータバッファの増設方法の一例について図１０のフローチャートを参照しながら説明する。なお、上位装置１００からのＩ／Ｏを受信しているディスクドライブ制御装置は、ディスクドライブ制御装置２００とする。

【００６３】データバッファの増設要求があった時、ステップ９１１にて当該ディスクドライブ制御装置はＩ／Ｏ受信側かを判断する。まずディスクドライブ制御装置２００の処理内容について説明する。ディスクドライブ制御装置２００はＩ／Ｏ受信側なので、ステップ９１２

に進み、ディスクドライブ制御装置４００がまず縮退し、切り放すことを認識する。そこで、データ転送制御部２３０にデータバッファ２４０へライトデータを１重書きとするよう指示をする。その後、ステップ９１３にて上位装置１００からのＩ／Ｏ処理を実行し、ステップ９１４にて、図５のステップ７００～ステップ７０３を実行する。すなわちディスクドライブ制御装置４００にて実行していた分を代替する。ステップ９１３とステップ９１４を繰り返しながらディスクドライブ制御装置４００の復旧完了を待つ。

【００６４】次にディスクドライブ制御装置４００もステップ９１１にて当該ディスクドライブ制御装置はＩ／Ｏ受信側かを判断する。その結果Ｉ／Ｏ受信側ではないので、ステップ９１５に進み、当該ディスクドライブ制御装置４００は切り放しを行ない、ステップ９１６にてデータバッファ４４０の増設を行なう。増設完了後、ステップ９１７にて復旧したことを通信機構３００を介してディスクドライブ制御装置２００に通知する。

【００６５】今度はディスクドライブ制御装置２００が増設を行なう必要があるため、ディスクドライブ制御装置４００はＩ／Ｏの受領を代替するため、ステップ９１９で切替機構４２０を用いてＩ／Ｏ受信するホストＩ／Ｆ制御部を自系に切り替える。その後、ステップ９２０にて上位装置１００からのＩ／Ｏ処理を実行し、ステップ９２１にて図５のステップ７００～ステップ７０３を実行し、ディスクドライブ制御装置２００の復旧完了を待つ。

【００６６】ステップ９１８で復旧を通信機構３００を介して認識したディスクドライブ制御装置２００は、ステップ９２２にて切り放し、ステップ９２３にてデータバッファ２４０の増設を行なう。増設完了後、ステップ９２４にて復旧を通信機構３００にてディスクドライブ制御装置４００に通知する。通知後、当該ディスクドライブ制御装置２００はホストＩ／Ｏ受信側ではないのでステップ９２５にて図５のステップ７００～ステップ７０５を実行する側に回る。

【００６７】ディスクドライブ制御装置４００は、ステップ９２６にて他系の復旧を通信機構３００にて認識すると、ステップ９２７にてデータ転送制御部２３０にデータバッファ２４０／４４０へライトデータを２重に書くよう指示をする。ステップ９２８にて上位装置１００からのＩ／Ｏ処理を実行する。

【００６８】このように上位装置１００からのＩ／Ｏを実行しながらも各系のデータバッファ２４０／４４０の増設が可能となる。すなわち本実施例によれば、従来ではデータバッファの増設はシステムを停止してからでないと実現できなかったのに対し、オンライン中に増設が可能となる。特に、低コストにて実現されている１つのボード上にディスクドライブ制御装置が構築されているときは、ボード毎の交換が必要なため、稼働中の増設は不可能であった。本実施例では、冗長構成のディスクドライブ制御装置２００／４００において、１台ずつを縮退／復旧させながら、データバッファの増設が可能である。

【００６９】また、本実施例によれば、図１０のステップ９１６および、ステップ９２３の処理をマイクロプログラム交換作業に置き換えることにより、稼働中のマイクロプログラムの交換が可能であり、２４時間運転の要求が著しい近年のコンピュータシステムにおける保守管理作業に特に有効である。

【００７０】また、片系障害時の縮退中、上位装置１００からのライト要求をデータバッファまでに書き込んで終了を報告するか、ディスク装置５００にまで書き込んで終了を報告するかはユーザが指示可能である。すなわち、このデータ転送モード情報３３０の書き換えは、ユーザのプログラムで自動的に行なってもよい。すなわち、データバッファが１面構成になった時、データファッバに格納した時点で終了を報告すれば、応答性にはすぐれているが、この時点でディスクドライブ制御装置に障害が発生すると、データ保証ができなくなる。一方、ディスク装置５００にまで格納するのでは、ライトペナルティ処理が発生してしまうため、応答性はかなり劣化してしまうが、上位装置１００に対しては、確実な応答が報告でき、信頼性は高い。本実施例の外部記憶装置の場合、ユーザが扱うファイルへの信頼度の要求レベルに応じて、ユーザの指示により、信頼度を優先するか、応答速度を優先するかを随意に選択でき、柔軟なファイルシステムを構築することが可能となる。

【００７１】さらに本発明では、複数のディスクドライブ制御装置は冗長な構成だけではなく、複数の上位装置または、複数のバスより、同時にアクセスが可能なシステムも提供できる。このシステム構成例を図１１および図１２に示す。

【００７２】図１１は、これまでに説明した実施例の図１と同じ構成だが、上位装置１００とのＩ／ＦがＳＣＳＩの時、図１の構成では記憶制御装置０と記憶制御装置１は同じＳＣＳＩＩＤで接続されていたのに対して、図１１の構成では記憶制御装置０（４００Ａ）と記憶制御装置１（２００Ａ）は異なるＳＣＳＩＩＤで接続されている点が異なっている。この図１１の構成の場合、どちらも、上位装置１００からＩ／Ｏ要求を受領して処理する。また、図１２は、複数の記憶制御装置０（４００Ｂ）および記憶制御装置１（２００Ｂ）が、同一の上位装置１００に対してマルチパスにより接続されたシステム構成の一例を示すブロック図である。この図１２の構成でも、記憶制御装置０（４００Ｂ）と記憶制御装置１（２００Ｂ）は、いずれも上位装置１００からのＩ／Ｏ要求を実行可能である。いずれがＩ／Ｏ要求を実行するかの指定は共通管理テーブル３１０の、ホストＩ／Ｏ受信情報３４０を書き換えることにより実現される。すな

わち、各記憶制御装置は、まずホストＩ／Ｏ受信情報３４０を参照し、当該記憶制御装置が上位装置からのＩ／Ｏを受信するか否かを決定する。このように、本発明では様々なユーザの接続方法に対応することができ、柔軟なシステムが構築できる。

【００７３】以上説明したように、本実施例によれば、冗長構成の複数のディスクドライブ制御装置２００／４００が負荷分散しながら上位装置１００からの要求を実行することにより、信頼性の向上だけでなく性能の向上も同時に実現することが可能なファイルシステムを提供できる。また、すべてのディスクドライブ制御装置２００／４００が負荷分散しながら上位装置１００からのＩ／Ｏ要求を実行しながらでも、障害発生時に上位装置１００からなんら指示を仰ぐことなく自動的に切り替わって稼働を継続し、さらに復旧することが可能となる。これにより、上位装置１００からのＩ／Ｏ要求を実行しながらデータバッファの増設やマイクロプログラムの交換が可能となり、無停止保守が実現できる。また、冗長構成だけでなく、すべてのディスクドライブ制御装置が同時に上位装置１００からの要求を受信する構成にすることも可能であり、ユーザの要求する多様なファイルシステムに柔軟に対応することができる。

【００７４】

【発明の効果】本発明の外部記憶装置によれば、冗長構成の複数の記憶制御装置に負荷を分散させることにより、信頼性および性能を向上させることができる、という効果が得られる。

【００７５】また、上位装置の側に記憶制御装置の冗長構成を意識させることなく、記憶制御装置の多重化による信頼性の向上、さらには記憶制御装置の多様な制御動作を実現することができる、という効果が得られる。

【００７６】また、稼働を停止させることなく、冗長構成の複数の記憶制御装置におけるハードウェアおよびソフトウェア等の保守管理作業を簡便に遂行することができる、という効果が得られる。

【００７７】また、単一のボード上に記憶制御装置およびデータバッファを搭載した構成の保守管理作業を、稼働中に実行することができる、という効果が得られる。

【図面の簡単な説明】

【図１】本発明の一実施例である外部記憶装置を含む計算機システムの一例を示す概念図である。

【図２】本発明の一実施例である外部記憶装置を構成するディスクドライブ制御装置の内部構成の一例を示すブロック図である。

【図３】本発明の一実施例である外部記憶装置において用いられる共通管理テーブルの構成の一例を示す概念図である。

【図４】本発明の一実施例である外部記憶装置の作用の一例を示すフローチャートである。

【図５】本発明の一実施例である外部記憶装置の作用の一例を示すフローチャートである。

【図６】本発明の一実施例である外部記憶装置の作用の一例を示すフローチャートである。

【図７】本発明の一実施例である外部記憶装置の作用の一例を示すフローチャートである。

【図８】本発明の一実施例である外部記憶装置の作用の一例を示すフローチャートである。

【図９】本発明の一実施例である外部記憶装置の作用の一例を示すフローチャートである。

【図１０】本発明の一実施例である外部記憶装置の作用の一例を示すフローチャートである。

【図１１】本発明の一実施例である外部記憶装置における上位装置との接続形態の変形例を示す概念図である。

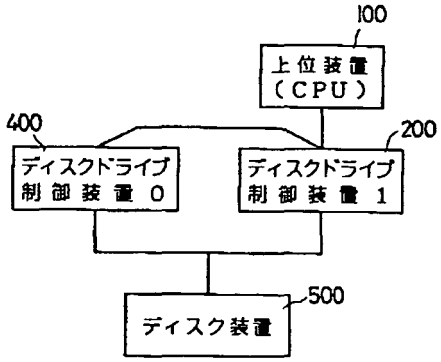
【図１２】本発明の一実施例である外部記憶装置における上位装置との接続形態の変形例を示す概念図である。

【符号の説明】

１００…上位装置、２００…ディスクドライブ制御装置、２１０…ホストＩ／Ｆ制御部、２２０…切替機構、２３０…データ転送制御部、２４０…データバッファ、２５０…マイクロプロセッサユニット、２６０…ＤＲＶ転送制御部、２７０…ＤＲＶＩ／Ｆ制御部、２８０…ＥＣＣ生成回路、３００…通信機構、３１０…共通管理テーブル、３２０…監視情報、３２１…監視情報、３２２…監視情報、３３０…データ転送モード情報、３４０…ホストＩ／Ｏ受信情報、３５０…負荷分散情報、４００…ディスクドライブ制御装置、４１０…ホストＩ／Ｆ制御部、４２０…切替機構、４３０…データ転送制御部、４４０…データバッファ、４５０…マイクロプロセッサユニット、４６０…ＤＲＶ転送制御部、４７０…ＤＲＶＩ／Ｆ制御部、４８０…ＥＣＣ生成回路、５００…ディスク装置。

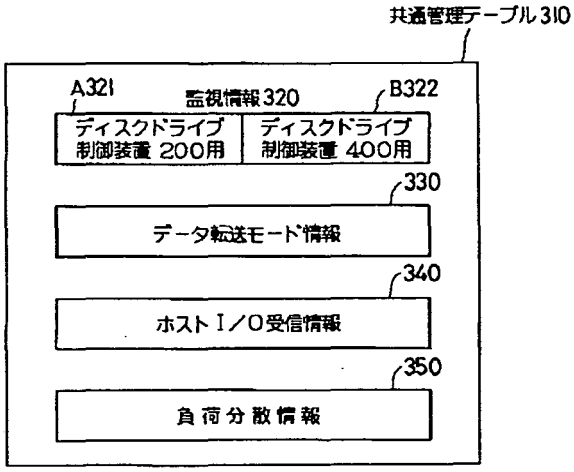
【図１】

図１



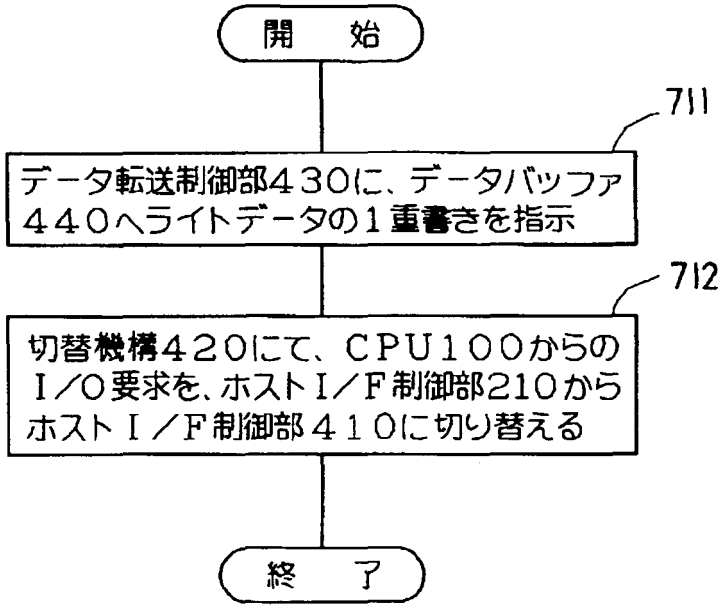
【図３】

図３



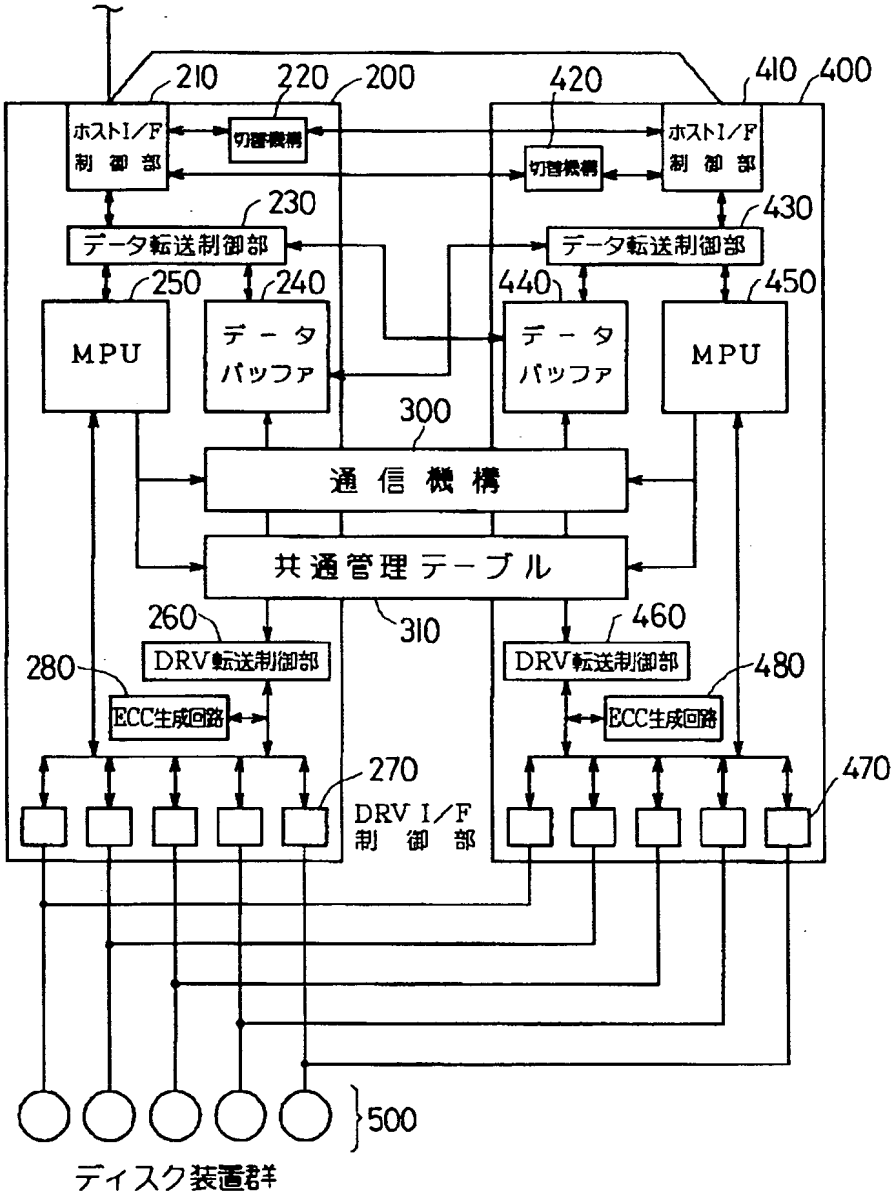
【図６】

図６



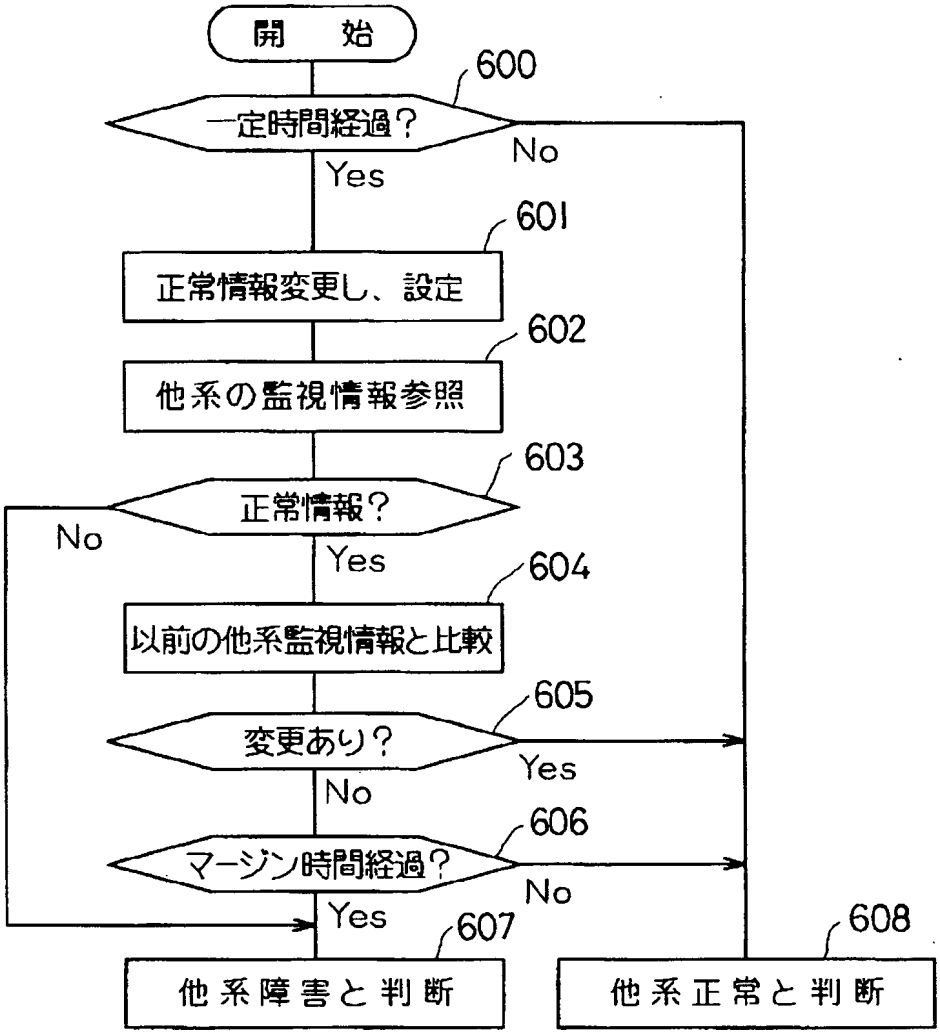
【図２】

図２


210
220
200
420
410
400
ホストI/F制御部
切替機構
切替機構
ホストI/F制御部
230
430
データ転送制御部
データ転送制御部
250
240
440
450
MPU
データバッファ
データバッファ
MPU
300
通信機構
共通管理テーブル
310
260
460
DRV転送制御部
DRV転送制御部
280
480
ECC生成回路
ECC生成回路
270
470
DRV I/F制御部
500
ディスク装置群

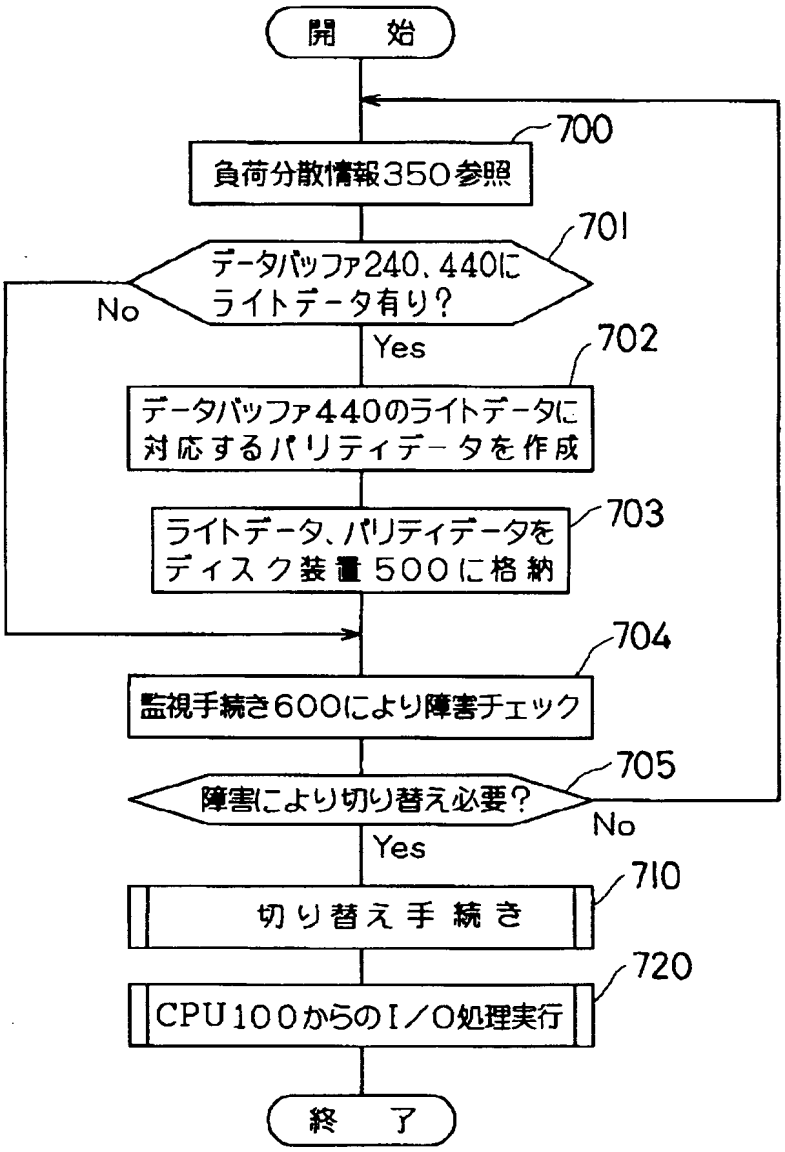
【図4】

図4


開　始
600
一定時間経過？
No
Yes
601
正常情報変更し、設定
602
他系の監視情報参照
603
正常情報？
No
Yes
604
以前の他系監視情報と比較
605
変更あり？
Yes
No
606
マージン時間経過？
No
Yes
607
他系障害と判断
608
他系正常と判断

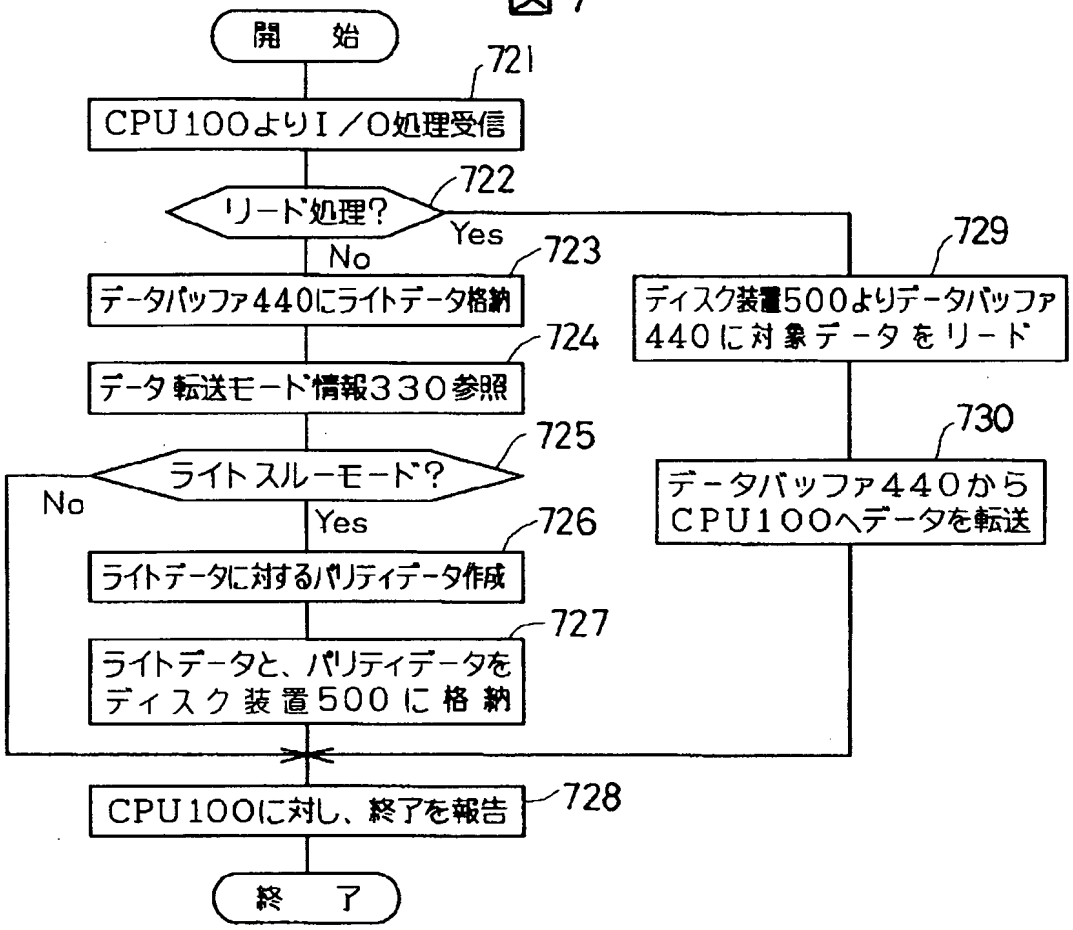
【図５】

図５


開　始
700
負荷分散情報350参照
701
データバッファ240、440に
ライトデータ有り？
No
Yes
702
データバッファ440のライトデータに
対応するパリティデータを作成
703
ライトデータ、パリティデータを
ディスク装置500に格納
704
監視手続き600により障害チェック
705
障害により切り替え必要？
No
Yes
710
切り替え手続き
720
CPU100からのI／O処理実行
終　了

【図７】

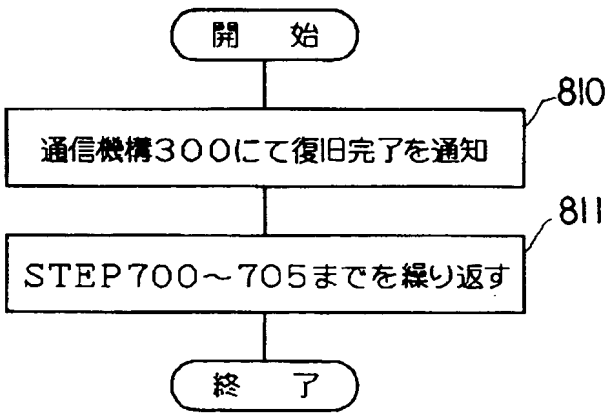
図７



【図８】

図８

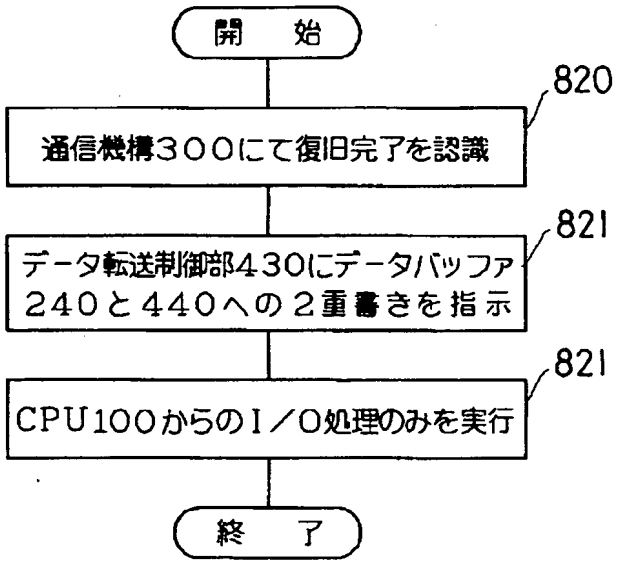
復旧方式（ディスクドライブ制御装置200）



【図９】

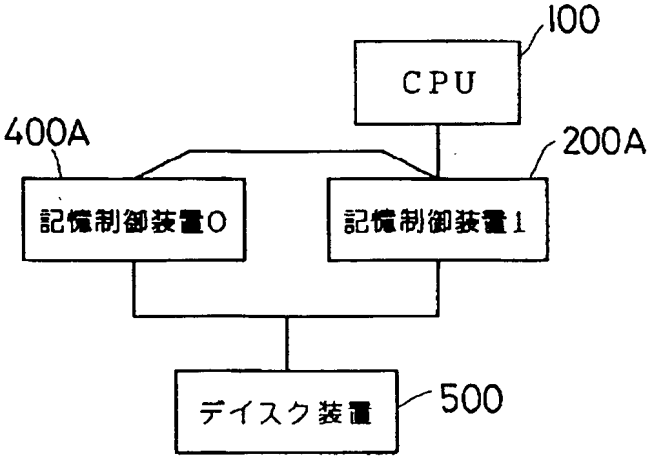
図９

復旧方式（ディスクドライブ制御装置４００）



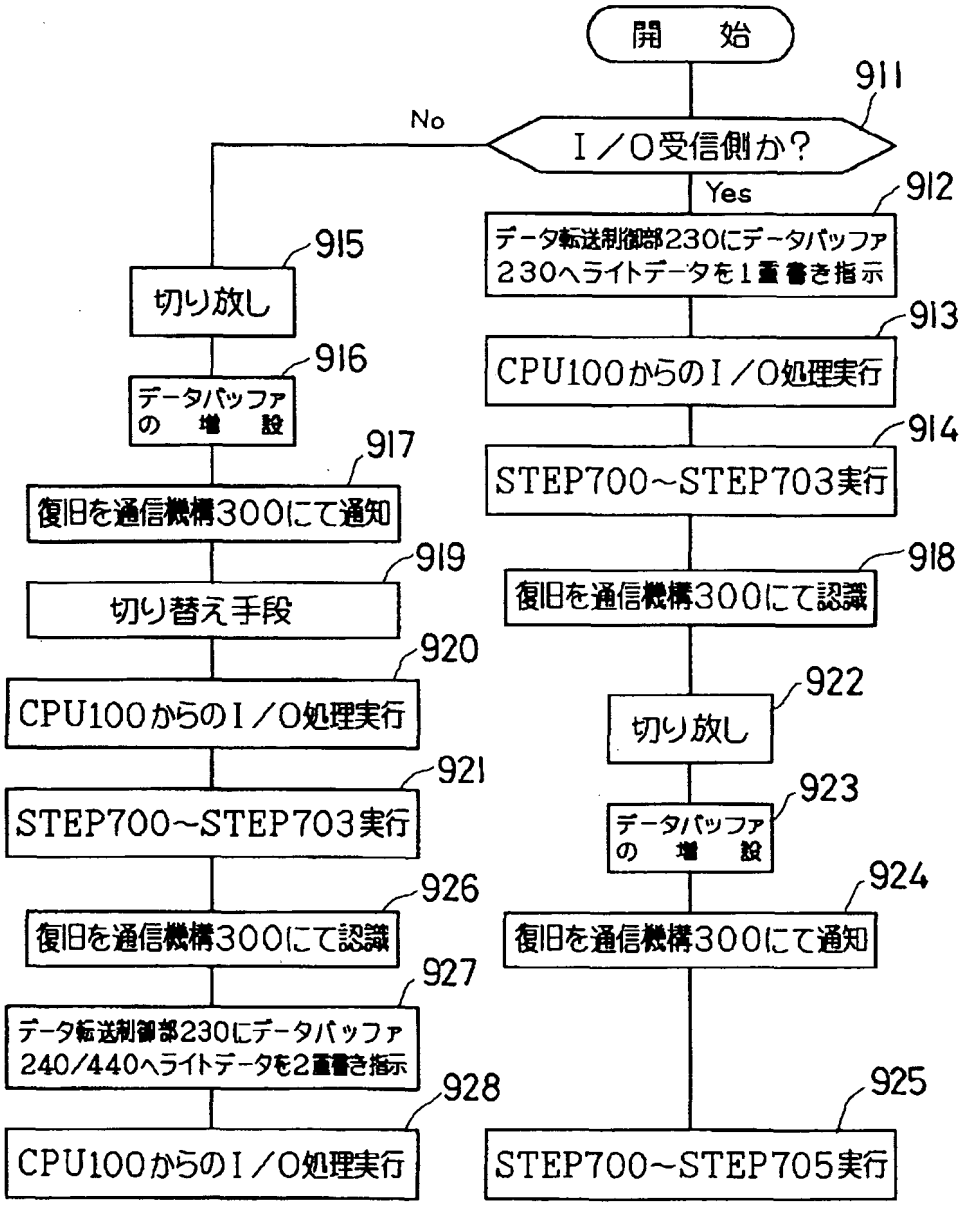
【図１１】

図11



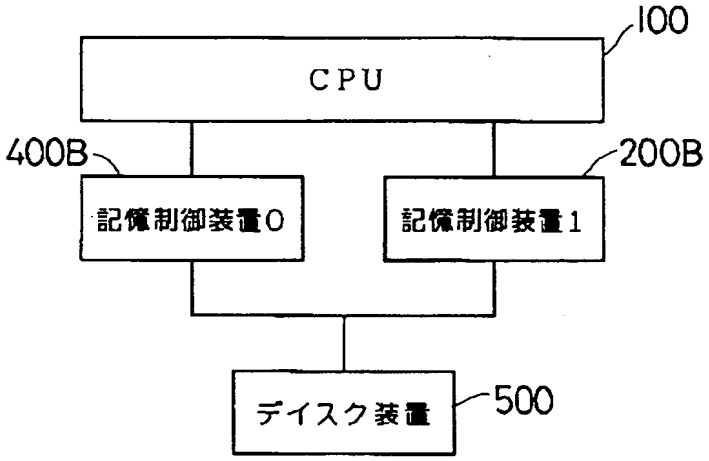
【図10】

図10


開　始
911
I／O受信側か？
No
Yes
912
データ転送制御部230にデータバッファ230へライトデータを1重書き指示
913
CPU100からのI／O処理実行
914
STEP700～STEP703実行
918
復旧を通信機構300にて認識
922
切り放し
923
データバッファの増設
924
復旧を通信機構300にて通知
925
STEP700～STEP705実行
915
切り放し
916
データバッファの増設
917
復旧を通信機構300にて通知
919
切り替え手段
920
CPU100からのI／O処理実行
921
STEP700～STEP703実行
926
復旧を通信機構300にて認識
927
データ転送制御部230にデータバッファ240/440へライトデータを2重書き指示
928
CPU100からのI／O処理実行

【図１２】

図12



【公報種別】特許法第１７条の２の規定による補正の掲載
【部門区分】第６部門第３区分
【発行日】平成１４年８月３０日（２００２．８．３０）

【公開番号】特開平８－３３５１４４
【公開日】平成８年１２月１７日（１９９６．１２．１７）
【年通号数】公開特許公報８－３３５２
【出願番号】特願平７－１３９７８１
【国際特許分類第７版】
G06F　3/06　304
　　11/20　310
　　12/16　310
　　13/14　310
【ＦＩ】
G06F　3/06　304 B
　　11/20　310 B
　　12/16　310 Q
　　13/14　310 F

【手続補正書】
【提出日】平成１４年６月６日（２００２．６．６）
【手続補正１】
【補正対象書類名】明細書
【補正対象項目名】特許請求の範囲
【補正方法】変更
【補正内容】
【特許請求の範囲】
【請求項１】　上位装置との間で授受されるデータが格納される記憶装置と、前記記憶装置と前記上位装置との間に介在し、前記上位装置と前記記憶装置との間における前記データの授受を制御する複数の記憶制御装置とを含む外部記憶装置であって、
複数の前記記憶制御装置が前記上位装置からみて等価に見えるように当該記憶制御装置を前記上位装置に接続するインターフェイス手段と、
個々の前記記憶制御装置に設けられ、他の前記記憶制御装置における障害または切替指令の有無を監視する監視手段と、
個々の前記記憶制御装置に設けられ、いずれの前記記憶制御装置が前記上位装置との間における前記データの授受の制御を行うかを切り替える切替手段と、
前記記憶制御装置の相互間における情報の伝達を行う情報伝達手段と、
前記上位装置からの入出力要求に起因する負荷を複数の前記記憶制御装置間にて分担させる負荷分散手段と、
を備えたことを特徴とする外部記憶装置。
【請求項２】　請求項１記載の外部記憶装置において、
複数の前記記憶制御装置の各々に設けられ、前記上位装置との間で授受される前記データを一時的に格納するデータバッファと、
前記上位装置からの書き込み要求時、複数の前記データバッファの各々に対して書き込み要求データを選択的または多重に書き込むとともに、前記書き込み要求データの前記データバッファに対する書き込み完了時点で前記上位装置に対して書き込み完了を報告し、前記上位装置からの入出力要求とは非同期に前記データバッファから前記記憶装置へ前記書き込み要求データを反映させるライトアフタ処理、および前記書き込み要求データの前記記憶装置に対する書き込み完了時点で前記上位装置に対して書き込み完了を報告するライトスルー処理を選択的に実行可能なデータ転送制御手段と、
を備えたことを特徴とする外部記憶装置。
【請求項３】　請求項１または２記載の外部記憶装置において、
複数の前記記憶制御装置から共通にアクセス可能にされ、個々の前記記憶制御装置が健全か否かを識別するための第１の管理情報、前記ライトアフタ処理および前記ライトスルー処理の何れを実行するかを指定する第２の管理情報、複数の前記記憶制御装置の何れが前記上位装置からの入出力要求を受け付けるかを指定する第３の管理情報、複数の前記記憶制御装置の各々における前記負荷の分担を指定する第４の管理情報の少なくとも一つが格納される管理情報記憶手段と、
障害の発生または外部からの切替指令を契機として、前記障害が発生したか、または外部から指令された前記記憶制御装置を切り離すとともに、残りの前記記憶制御装置によって前記上位装置との間における前記データの授受を継続する縮退運転を行う操作、および切り離されていた前記記憶制御装置を冗長構成に復帰させる操作を行う制御論理と、

を備えたことを特徴とする外部記憶装置。

【請求項４】　請求項２記載の外部記憶装置において、前記データ転送制御手段は、個々の前記記憶制御装置の各々に設けられた前記データバッファの各々に対する前記書き込み要求データの選択的な書き込み操作の停止および再開を行う制御論理を備えたことを特徴とする外部記憶装置。

【請求項５】　請求項４記載の外部記憶装置において、複数の前記記憶制御装置の中の少なくとも一つを選択的に停止させて縮退運転を行うとともに、停止された前記記憶制御装置に対応するデータバッファの保守または前記記憶制御装置を制御するマイクロプログラムの保守を実行することを特徴とする外部記憶装置。

【請求項６】　上位装置との間で授受されるデータが格納される記憶装置と、

前記記憶装置と前記上位装置との間に介在し、前記上位装置と前記記憶装置との間における前記データの授受を制御する第１の記憶制御装置と、

前記記憶装置と前記上位装置との間に介在し、前記上位装置と前記記憶装置との間における前記データの授受を制御する第２の記憶制御装置と、

前記記憶制御装置から共通にアクセス可能にされ、個々の前記記憶制御装置の管理情報を格納する管理情報記憶手段とを含む外部記憶装置であって、

通常は、前記第１の記憶制御装置が、前記上位装置と前記記憶装置との間で処理されるべき情報であって入出力要求を含むものを処理し、かつ、

前記上位装置と前記記憶装置との間で処理されるべき入出力要求に含まれる情報であって、前記記憶制御装置の各々における負荷の分担を指定する管理情報を、前記管理情報記憶手段に記憶することを特徴とする外部記憶装置。